VI. MODEL FIT AND CALIBRATION

It is critical that the estimated models effectively represent the data and serve as a valid tool for risk adjustment. Earlier sections of this report addressed issues of data validity and content validity. This section evaluates the discrimination and calibration of the CCMRP model.

Discrimination

Models that distinguish well between patients who die and those who survive are said to have good discrimination. A commonly used measure of discrimination is the c-index (also known as the c-statistic, or the area under the ROC curve). The c-index ranges from 0 to 1, with higher values indicating better discrimination. For the **1999** data model the c-index is 0.834. For the **All Quarters** data model the c-index is 0.816. In comparison, c-indexes reported in other recently published studies of CABG mortality using logistic regression (including those from New Jersey, New York, Pennsylvania, and the Society of Thoracic Surgeons) range from about 0.78 to 0.82. As such, the CCMRP models appear to discriminate as well as, or better than, those from other programs that produce risk-adjusted outcomes data for isolated CABG surgery.

Calibration

Calibration refers to the ability of a model to match predicted and observed death rates across the entire spread of the data. A model in which the numbers of observed deaths align well with the numbers of deaths predicted by the model demonstrates good calibration. Good calibration is essential for reliable risk adjustment. A common measure of calibration is Hosmer and Lemeshow's χ^2-statistic, which compares observed and predicted outcomes over deciles of risk. The Hosmer-Lemeshow test statistic is 28.9 (df=8; p-value=0.00) for the **1999** model and 29.0 (df=8, p-value=0.00) for the **All Quarters** model (i.e., reject the null hypothesis of no difference between actual and predicted deaths). This result was not a major cause for concern; with such a large sample it is common to fail the Hosmer-Lemeshow test.

The next step was to inspect the difference between the actual number of deaths and the predicted number of deaths (derived from the risk model) in each of the 10 groups. These groups are created by sorting all observations by the predicted risk of death and then dividing the sorted observations into ten groups of approximately equal size.

Tables 8 and 9 show the calibration of the **1999** and **All Quarters** risk-adjustment models.

Table 8: Calibration of 1999 Model

Group	N	Minimum Predicted Risk	Maximum Predicted Risk	Actual Deaths	Predicted Deaths	Difference
1	2,198	0.03	0.31	1	4.7	(3.7)
2	2,198	0.31	0.50	6	8.9	(2.9)
3	2,197	0.50	0.71	8	13.2	(5.2)
4	2,197	0.71	0.97	8	28.3	(10.3)
5	2,197	0.97	1.28	19	24.5	(5.5)
6	2,197	1.28	1.72	41	32.6	8.4
7	2,197	1.72	2.38	40	44.5	(4.5)
8	2,197	2.38	3.51	90	63.4	26.6
9	2,197	3.51	6.23	113	101.5	11.5
10	2,197	6.23	86.91	295	309.5	(14.5)

Table 8 provides a summary comparison of the **1999** model to the data. There are a total of 21,973 patients in the 1999 CCMRP dataset. The first row of the table represents the decile of patients at lowest risk of in-hospital death in the CCMRP model (i.e., the 2,198 patients whose predicted risk of dying ranged from 0.03 to 0.31%). Among the first decile, one patient died, but the model predicted death for five of the patients. Assuming a Poisson distribution for a binary outcome with mean 0.0023 (5 ÷ 2,198), the predicted range of deaths for the first decile is 0.6 to 9.3 deaths. Thus, the one death that occurred falls within the expected range.

The last row of Table 8 represents the highest risk decile of patients for 1999. Among this group, 295 died whereas the model predicted 310 deaths. The predicted range for the tenth decile is 275 to 345 deaths. Again, the number of observed deaths falls within the expected range.

Table 9: Calibration of All Quarters Model

Group	N	Minimum Predicted Risk	Maximum Predicted Risk	Actual Deaths	Predicted Deaths	Difference
1	4,983	0.05	0.39	7	13.8	(6.8)
2	4,983	0.39	0.58	14	24.1	(10.1)
3	4,983	0.58	0.78	22	33.8	(11.8)
4	4,982	0.78	1.02	34	44.7	(10.7)
5	4,982	1.02	1.31	64	57.2	6.2
6	4,982	1.31	1.69	71	74.3	(3.3)
7	4,982	1.69	2.28	102	97.7	4.3
8	4,982	2.28	3.26	143	135	8.0
9	4,982	3.26	5.56	251	209.7	41.3
10	4,982	5.56	88.43	608	626.1	(17.9)

Figures 5 and 6 contain additional representations of the model calibration. The left panel of the graph plots the cumulative number of *predicted* deaths against the number of *actual* deaths. The closer the predictions are to the actual number of deaths, the closer the curve is to the superimposed 45-degree line. Overall, the predictions appear to track the actual observed deaths well.

The right panel plots the Actual and Predicted number of cumulative deaths against all **1999** and **All Quarters** cases respectively. The “smooth” curve summarizes the model predictions, while the slightly jagged curve represents the actual deaths. Because the models calibrate to the data well, the two curves lie close to each other. In addition, both curves are relatively flat on the left side and increase rapidly as they move toward the right, akin to so-called “exponential” curves. This suggests that the majority of CABG surgeries are low in risk and that most in-hospital deaths occur in higher-risk patients. For 1999, only 42 deaths occurred among the 10,988 patients that fall in the lower half of the risk profile. Conversely, the remaining 579 deaths are concentrated in the 10,985 cases in the upper half. Although the overall in-hospital mortality rate following isolated CABG surgery is only 2.83%, the average risk of death for those in the lower half of the risk profile is 0.6% as opposed to 5% in the upper half. Although the graph does not show it, a straight line connecting the lower leftmost point with the upper rightmost point identifies a “constant risk” line of 2.83%, and would serve to demonstrate the importance of risk adjusting CABG data.

Three features stand out about the calibration of the model:

- The majority of cases exhibit low risk. Nonetheless, the range of predicted risks (from almost zero to almost 90%) seems adequately wide, suggesting that the model does cover the entire range of risk levels. This should ease concerns that the risk models cannot adequately adjust for high-risk patients.
- The model fits quite well in the higher risk categories. For 1999, among patients whose predicted risk exceeds 6.23%, the number of predicted deaths (408) approximates the number of observed deaths (412). This suggests that the risk-adjustment model works quite well for higher risk patients. As such, the model does not provide an incentive for hospitals to exclude high-risk patients from appropriate surgeries in order to improve their risk-adjusted rates.
- There is evidence that the model over-adjusts at the lowest risks, but this evidence is not statistically significant and the over-adjustment is relatively small.[18]

[18] A test was performed to determine whether the expected range covers the number of observed events in each group. Only for group 8 in the 1999 analysis is the observed count of deaths outside the range of the expected deaths.

Figure 5: Calibration of 1999 Model

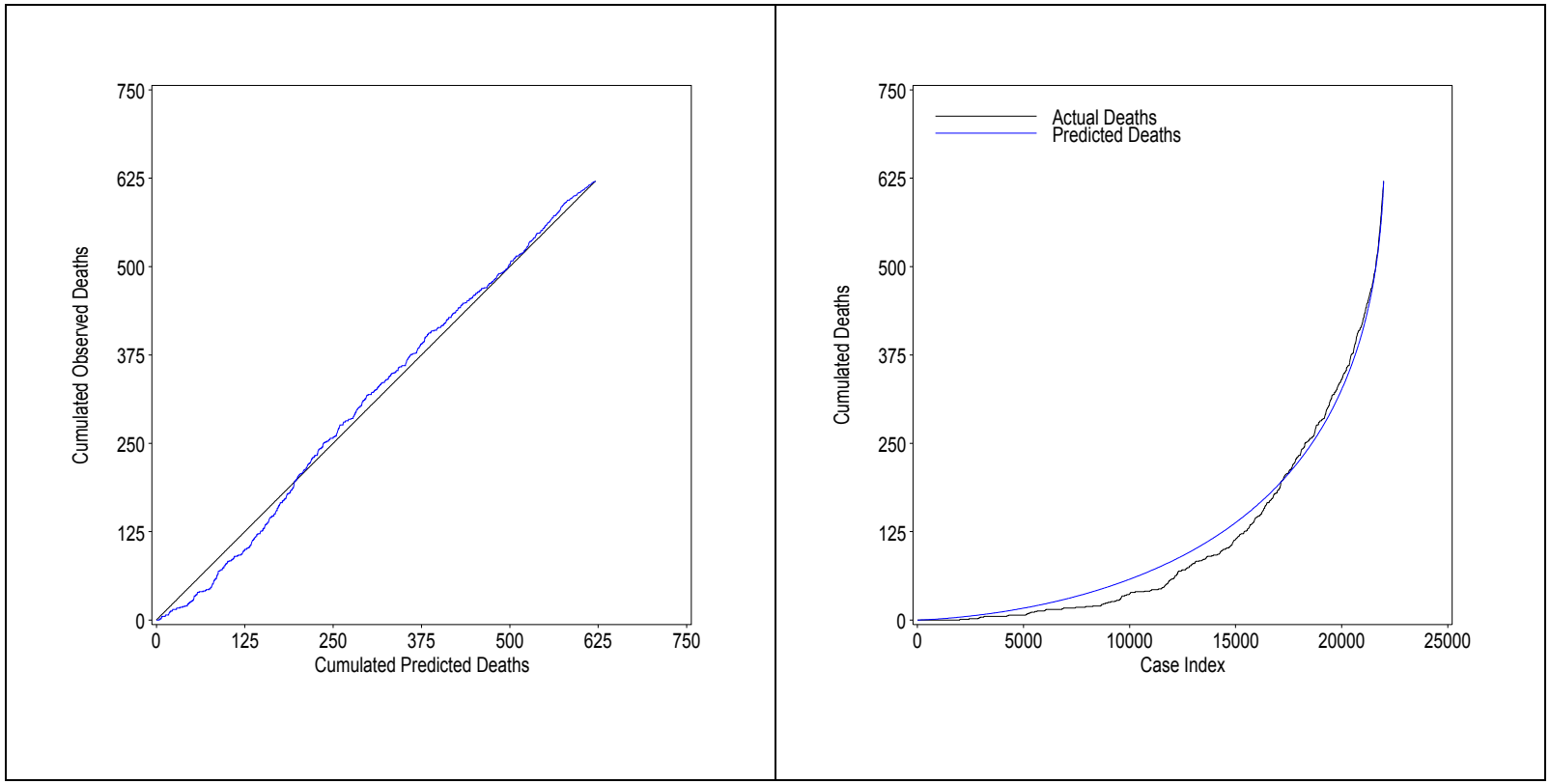

Figure 6: Calibration of All Quarters Model